Available online at www.sciencedirect.com
ScienceDirect
Procedia Economics and Finance 15 (2014) 1603 – 1612
Procedia
Economics and Finance
www.elsevier.com/locate/procedia

Emerging Markets Queries in Finance and Business

Assembling international equity datasets – review of studies on the cross-section of returns

Antonina Waszczuk[a,*]

[a]*European University Viadrina Frankfurt (Oder), Große Scharrnstraße 59, 15023 Frankfurt (Oder), Germany.*

Abstract

This paper discusses the data sources used in the international research on the cross-section of stock returns. Covering the wide range of internationally focused papers I give the overview of the applied data, sample coverage, classification schemes and data cleaning methods. I address the quality concerns in case of the non-U.S. data and methodologically relevant specifics of international data analysis providing references to available solutions. In regards to data cleaning I give an overview of applied screens, pointing out their diversity across studies. On that way I offer the first structured insight into challenges and specifics of rapidly increasing amount of papers discussing the cross-section of common stocks in both single-country and multiple-country frameworks.

Selection and peer-review under responsibility of the Emerging Markets Queries in Finance and Business local organization

Keywords: cross-section of stocks; data quality; data cleaning; empirical research; data errors

1. Motivation

Up to recently, dominant number of papers on the cross-section of stock returns has been conducted using the U.S. equity data, making it the academically best explored market in the world. The advantages of the U.S. data, apart from the role and importance of the U.S. financial sector, are the length of the time-series, number of listed stocks and data quality. However, the common base for majority of empirical papers causes the problem labelled as data snooping, i.e., the concern *"that prior empirical research may influence the way current investigations are conducted"*, Lo and MacKinlay (1990). Consequently, the statistical testing might support incorrect statements if results are not considered in the context of past inferences, see also Kothari et al. (1995), MacKinlay (1995).

Tests on the persistence of patterns found in the cross-section of the U.S. equity data present a possibility to discard such data-related criticism. Relevant approach requires either (a) use of an alternative methodology or (b) use of an alternative data population. The first proposition refers to empirical design given that studies on the cross-section are subject to many arbitrary choices in regards to criteria like data frequency, weighting and classification schemes, data filtering or length of the estimation period. Second proposition gears towards the variation of input

* Corresponding author. Tel.: +49-179-652-0408
E-mail address: euv36052@europa-uni.de

Selection and peer-review under responsibility of the Emerging Markets Queries in Finance and Business local organization
doi:10.1016/S2212-5671(14)00631-5

data, e.g. consideration of international markets as the source of alternative datasets or manipulation of the time-frame of the study. Over the long time the important problem of the international equity research was the difficulty in assembling a comprehensive international dataset constituting the individual stock data. Recently, the availability of the international data has improved. Further, over the last twenty years new markets arose, e.g., emerging Europe and Chinese stock exchanges in Shanghai and Shenzhen that were opened in early 1990s. Although the length of the data time-series is significantly shorter than for the U.S. market, they present an interesting alternative data population. Their attractiveness stems from their (partial) segmentation from developed capital markets and deviating qualitative market characteristics like market infra- and microstructure or market efficiency and influence of different political and economical regimes what might affect the risk-return profile of assets. Further international data are possibly free from database specific biases and cover the market specific characteristics offering the opportunity to test relevant hypothesis under different market conditions. Over the years, empirical research has accumulated large amount of papers delivering mixed results implying the differences across international equity markets. A notable recent example is delivered by Chui et al. (2010) who show that momentum profits are significantly related to some of country-specific variables and absent in the countries with the low level of individualism.

This paper gives an extensive overview of international data sources and related quality issues and way to mitigate them. It provides numerous references to international empirical papers considering both single-country and multiple-country. To keep the focus I concentrate on three most prominent patterns related to size, value and momentum. The rest of the paper is structured as follows. Section 2 gives overview of data sources used in both locally and internationally focused papers. Section 3 discusses the coverage of market. Biases present in the databases are outlined in Section 4. Section 5 concludes the paper.

2. Data sources

2.1. U.S. data

The primer source of U.S. stock data on the security prices, returns, and volume is the Center for Research in Securities Prices (CRSP). The data from the CRSP database have been used or referred to in the one-third of the studies in empirical finance since available, see Economist (2010). CRSP provides time-series going back to 1926 for the NYSE, 1962 for Amex and 1972 for Nasdaq. The launch of the CRSP database in 1960s and initial use of the data is reviewed by Weinstein (2010) and Fama (2011). The starting date for the time series set 1926 results from the intention of the founders to capture at least one whole business cycle of the NYSE history. Goetzmann et al. (2001) discuss the data available for the pre-1926 period but this pre-CRSP sample is not commonly investigated, mainly due to its quality.

The source of the U.S. firms' accounting data is the Compustat that provides data series back to 1962 on quarterly basis and back to 1950 on annual basis. Davis et al. (2000) supplement the Compustat book common equity data of industrial firms that do not have Compustat data with the data from Moody's Industrial Manuals to extend the coverage of the first years of the sample period. This data set going back to 1920s represents the most extensive accounting data sample for the U.S. market and underlies the standardised risk factors in the K. French data library ().

Most of the cross-sectional analysis on the U.S. data requires use of both CRSP and Compustat. Initially, matching the data from both sources caused series of problems related to inconsistent and changing Cusip numbers, Chan et al. (1995). Recently, the merged sample is available as the CRSP/Compustat Merged Database (CCM) and perceived as a high quality base for the research of U.S. equity markets. Nevertheless, it is not free from errors as discussed in Rosenberg and Houglet (1974), Ball and Watts (1979), McElreath and Wiggins (1984), Courtenay and Keller (1994), Bennin (1980) and Ince and Porter (2006), among others.

2.2. International data

Recently, for non–U.S. data, the Thomson Reuters DataStream (TDS) is the regularly used source offering the most comprehensive set in terms of the covered markets (over 175 countries) and the number of securities per market (57.000 companies). Bekaert et al. (2007) state that the total number of firms per market available through TDS accounts for, on average, about 90% of domestically listed firms reported by the World Bank's World Development

Indicators. The TDS itself claims to cover 95% of global market value with up to 20 years of historical data. De Moor and Sercu (2013) argue that for many countries TDS contains information about all listed stocks. TDS contains data from two databases: returns and market capitalization data come from the DataStream while data like fiscal year endings, book value and market value on the fiscal year ending come from the Worldscope. TDS is the data source underlying the papers by Griffin (2002), Brown et al. (2008), Nijman et al. (2004), Chui et al. (2010), Guo and Savickas (2010), Narteaa et al. (2011), Hou et al. (2011), De Moor and Sercu (2013) and Cakici et al. (2013), among others.

Raising popularity and coverage of the TDS made it also to an alternative source of the U.S. data enabling a cross-check with the CCM data. The increasing number of publications deals with the quality of the equity data stating that any use of pricing models and risk factors relies on the credibility of information available. Frequently cited paper by Ince and Porter (2006) shows that momentum effect is not present in the raw U.S. data from TDS. The authors present two levels of screens and filters necessary to clean the data from TDS to make both U.S. data sources comparable and the TDS data suitable for further analysis. The extend to which these screening procedures, discussed in details in the paper, are considered by scholars working with TDS data is unknown. Several studies follow the recommendations, e.g., Naranjo and Porter (2010), De Moor and Sercu (2013) or Hou et al. (2011). The latter paper conduct the study using both CCM and TDS data for the U.S. market and shows the robust evidence. Guo and Savickas (2010) apply their own screens that seem to be similar to those by Ince and Porter (2006) without referencing them. Other studies comparing both CRSP/Compustat and TDS data environments are, e.g., Schmidt et al. (2011) and Ulbricht and Weiner (2005). Chui et al. (2010) carry out a cross-check between TDS and PACAP for Asian Markets and confirm the consistency between both. On the other hand, Brückner (2013) presents several sources of data mismatches in the TDS data for the German market and points out the problems relevant for the final outcome of studies. Rossi (2011) discusses the data errors in TDS relevant for the UK market. Similar studies do not exist for other single markets what makes a general assessment about the influence of TDS data quality on the empirical analysis impossible at this stage.

Besides TDS, several other data vendors are in use when building an internationally scoped dataset. Fama and French (1998) and Bauer et al. (2008) use the Morgan Stanley's Capital International Perspectives (MSCI) when examining developed markets. Bauman et al. (1998) uses Compustat Global Vantage file. The PACAP from the Pacific-Basin Capital Markets Research Center at the University of Rhode Island represents an alternative data source for the Asian markets used by e.g., McInish et al. (2008) and Chui et al. (2010). Studies investigating pricing mechanisms on the emerging markets use the Standard & Poor's International Finance Corporation's (S&P/IFC) Emerging Markets Database (EMDB), e.g., van der Hart et al. (2003), van der Hart et al. (2005), Bekaert et al. (2007), Umutlu et al. (2010), Borys and Zemcik (2011). To investigate the frontier markets De Groot et al. (2012b) use Standard & Poor's Frontier Board Market Index (S&P/FBMI) and collect the data on total returns from the Interactive Data Exshare. Alternative source of return data for both developed and emerging markets is the Factset Pricing used by, e.g., Heston et al. (1999), Nijman et al. (2004) (for price data), van der Hart et al. (2005), Umutlu et al. (2010).

As stated by Gregory et al. (2013), combining several data sources enables to infill any missing data in TDS and raise the quality of data sample. Indeed, international samples are often biased towards large stocks and are therefore not fully representative. To ensure the high quality of the datasets, many studies supplement the primary data source with the data from other databases. Daniel et al. (2001) merge PACAP database with Daiwa Securities and Nihon Keizai Shimbun files to precisely cover the Japanese sample. van der Hart et al. (2005) compare total and price returns from S&P/IFC with the Factset Pricing database to ensure the quality of return observations. Bekaert et al. (2007) uses monthly returns from EMDB and daily from TDS. To investigate the possibly wide spectrum of stocks including the micro-caps, Fama and French (2012) take the international stock returns and accounting data from Bloomberg and supplement them by TDS. De Groot et al. (2012b) works with problematic sample of frontier markets that suffer significant data-quality problems and therefore cross-check the data from S&P FBMI with Interactive Data Exshare, Bloomberg, TDS and data from the local stock exchanges. Chui et al. (2003) uses NEEDS, PACAP and TDS databases for different subperiods of investigated time window to guarantee the highest data quality and coverage.

2.3. National data

Publications on the cross-section of stock returns have been dominated by analysis of multiple country sets, mainly because in pooled samples more stocks can be used what raise the quality of statistical analysis. On the other hand,

however, inferences about particular country are hardly to draw from such data. Also, market integration might not be a reasonable assumption for all considered country constelations. For those reasons, studies on single countries provide a necessary supplement for understanding the cross-section of stocks around the world.

Further advantage of one country analysis is that the authors make often use of national data delivered by local stock exchanges or data providers. As noted by Schmidt et al. (2011), these data might be inaccessible to other researchers. Also the quality of the data is easier to assemble on a national level with the degree that is often unmatched by the public data providers. Such independent datasets give an opportunity for the out-of-sample tests with high coverage of the market and freedom of possible biases, see, e.g., Nagel (2001). At the same time, however, the comparability of results might cause some difficulties due to the possible discrepancies in variable definitions and coverage of the sample.

The examples of the alternative data sources for the individual non-U.S. markets are:

- Australia: Securities Industry Research Center of Asia Pacific (SIRCA), Drew et al. (2006) or Australian Graduate School of Management database (AGSM) for market variables and Aspect Financial database for accounting data, Dempsey (2010),
- Canada: financial statement data from the Financial Post database and from Research Insight Compustat and the market data from Toronto Stock Exchange-Western tape supplemented by Research Insight Compustat, LHer et al. (2004),
- China: Taiwan Economic Journal's (TEJ) China database, Eun and Huang (2007) or China Stock Market and Accounting Research (CSMAR), Huang et al. (2012),
- Germany: Karlsruher Kapitalmarktdatenbank (KKMDB) in Karlsruhe for the stock prices and Saling/Hoppenstedt Aktienführer for accounting data, see Artmann et al. (2012). Alternative database is maintained at Humboldt University in Berlin by Richard Stehle, Schulz and Stehle (2002) and Brückner et al. (2012).
- Japan: Pacific-Basin Capital Markets database, Chan et al. (1998), Daniel et al. (2001) or Griffin (2002),
- Poland: Bulletins of the Warsaw Stock Exchange, Lischewski and Voronkova (2012) and Waszczuk (2013),
- South Korea: database of Korea Capital Market Institute and Kis-Value, Kim et al. (2012),
- Switzerland: Factset, Ammann and Steiner (2008),
- United Kingdom: London Business School Share Price Database, Fletcher (2001), Lee et al. (2007), Gregory et al. (2009), Gregory et al. (2013).

Besides above examples, TDS is often used as a source of the single-country data, see Burghof and Prothmann (2011) and Siganos (2010) for UK, Griffin (2002) for UK and Canada, Glaser and Weber (2003), Amel-Zadeh (2011), Bank et al. (2012) and Hanauer et al. (forthcoming) for Germany, Akdeniz et al. (2000) for Turkey or Diacogiannis and Kyriazis (2007) for Greece.

3. Market and country coverage

Most of the recent studies on the U.S. equity market work with the universe of NYSE, Nasdaq and Amex stocks. Some studies use only the NYSE data argumenting either that the NYSE is representative for the U.S. equity market covering the majority of capitalisation or that bias can be introduced due to the heterogenous market structure. Nasdaq and Amex are known to be dominated by smaller stocks and the exclusion of those two exchanges decreases the strength of the size effects in the sample. Therefore the results obtained basing on the NYSE environment do not allow to draw much conclusions regarding the size-related effects.

Literature has recognised also the phenomenon labelled the "Nasdaq effect". Reinganum (1990) shows that returns of NYSE securities are about 6% higher than returns of securities listed on Nasdaq during the period 1973-1988. Loughran (1993) attributes most of Nasdaq stocks' underperformance to the underperformance of IPOs which is proportionately more important on Nasdaq. Fama and French (1993) find that the difference between NYSE and Nasdaq returns for size sorted portfolios is not significant after risk adjustment by the Fama-French factors. Loughran (1993) and Brennan et al. (1998) show the opposite. To account for possible influence and differences, some studies introduce Nasdaq dummy variable, e.g., Brennan et al. (1998). Other type of robustness test is to re-estimate the results for the Nasdaq-only subsample. Following this method, Gutierrez and Kelley (2008) show no weekly reversal in returns, Bulkley and Nawosah (2009) no momentum and Goyenko et al. (2009) no liquidity effect in the Nasdaq market. Liu (2006) carries out a comparative study for NYSE/Amex/Nasdaq versus NYSE/Amex universe and, alternatively,

examines two subsamples, NYSE/Amex and Nasdaq stocks separately. Kothari et al. (1995) and Yao (2012) run their analysis only for NYSE and Amex although the latter paper mentions the robustness of the results after Nasdaq inclusion.

Studies on non-U.S. equity markets include mostly the largest stock exchange in the country measured, e.g., in terms of number of stocks, i.e., Tokyo Stock Exchange, London Stock Exchange, Toronto Stock Exchange, see Daniel et al. (2001), Griffin (2002), Naranjo and Porter (2010), Chui et al. (2010). Argument supporting such choice is that these exchanges cover majority of relevant market capitalisation. Hou et al. (2011) and Huang et al. (2012) consider more than one exchange for large financial centers like China or Japan.

Many exchanges are divided into several segments targeting firms with different capitalisation, requirements regarding the transaprency and reporting or availability to foreign-investors. Japanese stock exchange covers three different segments, First, Second and Mothers Sections. Chinese A-share class is tradeable only for domestic investors. German Frankfurter Stock Exchange have Top Segment and

Further, international studies mostly focus on a particular subset of countries. The choice of the sample set might follow one of the listed criteria.

- level of markets development: developed markets (Heston et al. (1999), Fama and French (1998), Fama and French (2012)), emerging markets (Achour et al. (1998), Rouwenhorst (1999), Cakici et al. (2013)) and frontier markets (De Groot et al. (2012a)),
- the composition of official indices, e.g., constituents of the the S&P Frontier Board Market Index (De Groot et al. (2012b)), the Morgan Stanley Developed Country Index (Bekaert et al. (2012)),
- size and regional importance: U.S., UK, Japan (Chan et al. (1998), Daniel et al. (2001), Zhang (2006)),
- pre-defined country sets: G7 (Guo and Savickas (2010), Eiling et al. (2012)) or European Monetary Union (Ammann et al. (2012)),
- geographical location: global regions (Fama and French (2012), Cakici et al. (2013)), European countries (Bauer et al. (2008)), Visegrad countries (Borys and Zemcik (2011)), Asian or Southeast Asian stock markets (Brown et al. (2008) and Narteaa et al. (2011), respectively), Gulf Cooperation Council (Bley and Saad (2012)) or West African Economic and Monetary Union (Soumare et al. (2012)).

Analysis of multiple-country sets can be conducted on both pooled and country-by-country level. In case of aggregate approach the results might be driven by country effects, i.e., by the results for the one or several countries, see Rouwenhorst (1999) and Ammann et al. (2012). Brown et al. (2008) argue that inclusion of Taiwan into their basket of investigated countries affects the large picture because, unlike remaining countries in the sample, Taiwan exhibits value discount.

4. Biases in equity databases

Data samples obtained from the data providers discussed above are subject to several biases that might influence the outcome of the empirical analysis. Below, I discuss the most important examples.

Sample selection bias. Selection bias is a consequence of the selection rule other than random sampling that causes some observations to be excluded from the sample, e.g., due to the data availability. The sample is also selection-biased when it has some specific characteristic non-existent in alternative samples. It is the case in Daniel et al. (2001) who explore Japanese equity market characterised by high book-to-market premium relatively to other markets. Drew2008 talks about concentration of firms within few sectors for Australian market. As stated by Rouwenhorst (1999), SP/IFC EMDB uses special criteria to include stocks into their indices in order to reflect the local market's best and therefore the database is biased towards larger, more frequently traded stocks. MSCI database includes only large stocks constituting 80% of the market capitalisation for twelve developed markets investigated by Fama and French (1998). The SP/IFC global index aims to represent 70-80% of the total market capitalization of the local stock exchange. SP/FBMI covers around 80% of the capitalisation. As so, the international samples are usually not fully representative and some effects, mainly the size effect, cannot be investigated, see Kothari et al. (1995) for the U.S. and Fama and French (1998) for the international data consultation. For that reason Nijman et al. (2004) observe no clear size and value effects in their sample which covers fewer firms than sample by Heston et al. (1999) extracted from the same database for which the size effect for European stocks is restricted to the smallest three deciles.

With time new and small firms have become less under-represented in the TDS. As a consequence, results from studies using this data source might differ across time intervals due to changes in the coverage of the database. For example Brückner (2013) discusses the coverage issues in the TDS prior to 1990 that are relevant for the German market.

Survivorship bias. Survival bias is an example of the sample selection bias driven by the disproportionate exclusion of stocks that were delisted over time. Kothari et al. (1995), among others, argue that the inclusion of stocks that were distressed and survived is more likely in Compustat than inclusion of data on stocks that died. Survivors are likely to have unexpectedly high returns in the years just before inclusion in the Compustat and the database tends to back-fill the accounting data for small firms which were subsequently extreme winners, see Fama and French (1996). Such sample is biased towards winners because it contains mostly stock with better performance. Therefore, Jagannathan and Wang (1996) claim that Compustat is unsuitable for econometric analysis of asset pricing models. Wang (2000) demonstrate that survival alone can cause the size effect and book-to-market effect. However, followed up on the arguments and evidence, many studies obtained results consistent with survivorship bias, e.g., Davis (1996) who note however that the significance of estimates from Compustat alone may cause overstated coefficients. Fama and French (1992), Kothari et al. (1995) and Avramov and Chordia (2006) control for the Compustat survival bias by excluding the first two years of Compustat data for every firm. Horowitz et al. (2000) avoid the survival bias just because the authors does not include accounting data from Compustat when investigating the isolated size effect. For more details on the survivor bias in Compustat and for further references, see, e.g., Chan et al. (1995) and Fama and French (1996).

For majority of countries TDS does not include delisted securities prior to 1991, so the data prior to this date suffer from survivorship bias, Griffin (2002). PACAP does not include delisted stocks prior to 1988. However, Daniel et al. (2001) argue that the delisting rate is less than seven stocks a year and use of value-weighted portfolios can effectively mitigate the bias. Rouwenhorst (1999) and van der Hart et al. (2003) do not include stocks imidiately after their data are available in the S&P/IFC EMDB database but only after they are included in the IFC Composite Index so the backfilling bias is controlled for. Bauer et al. (2008) claim that MSCI database is not affected by the survivorship bias.

Another source of the survivorship bias emerges from the fact that the TDS exchange information often reflects only the current value of the classification variables, Ince and Porter (2006). As a consequence, only the most successful stocks that remain in the major exchanges are taken into account. Brückner (2013) discusses this problem for the case when only TDS data from the Top Segment of the German market is analysed.

Delisting bias. Several studies show that the CRSP database omits delisting returns for a large number of companies for the month in which a company is delisted from an exchange. Shumway (1997) notes that the omitted delisting returns, that average around -30%, can have important consequences for research applications. Shumway and Warther (1999) investigate this issue for the stocks listed on the Nasdaq exchange. The authors collect the data on delisting returns from the over-the-counter (OTC) market and replace the missing observations for delisted stocks with the proposed -55% return. After such correction they do not document the size effect on the Nasdaq exchange. On the other hand, Eisdorfer (2008) shows that on average about 40% of the momentum profit is generated by the returns of stocks that are delisted from the market during the holding period.

The extend to which the delisting bias should be controlled for depends on the amount of the surprisingly delisted stocks. Given that for many exchanges the stocks are typically not delisted unexpectedly, this concern might be less important than for the U.S. market. De Moor and Sercu (2013) show delisting bias to be irrelevant for their international data sample. Also Nijman et al. (2004) argue that missing returns adjustment for bankrupcy firms has limited influence on their momentum study and that results are more conservative because losers returns, i.e., the shorted side, to which the bankrupt firms and their highly negative final returns would be classified, are even more negative than expected. However, the relevance of the delisting bias for the non-U.S. data samples is rarely raised in the empirical papers.

5. Conclusions and Remarks

The ongoing internationalisation of capital markets is visible also in the recent academic publications investigating the cross-section of stock returns. The raising interest in universal validity of relationships documented for the U.S.

equity market as well as profits from the international diversification result in numerous studies focused on broader universe of markets. Analysis of international datasets shares many data-related problems with the analogous analysis of the U.S. market. But it also faces several challenges unique for the multi-country datasets. This paper presents in a structured way the variety of data sources and coverage of the data used in the international research on equity markets. Large amount of references to relevant papers serves to provide useful insight into the current focus of the research on the cross-section on the international equity markets.

References

D. Achour, C. Harvey, G. Hopkins, and C. Lang. Stock selection in emerging markets: portfolio strategies in Malaysia, Mexico, and South Africa. *Emerging Markets Quarterly*, 2:38–91, 1998.

L. Akdeniz, A. Altay-Salih, and K. Aydogan. A cross-section of expected stock returns on the istanbul stock exchange. *Russian and East European Finance and Trade*, 36:6–26, 2000.

A. Amel-Zadeh. The return of the size anomaly: Evidence from the German stock market. *European Financial Management*, 17:145–182, 2011.

M. Ammann and M. Steiner. Risk factors for the Swiss stock market. *Swiss Journal of Economics and Statistics*, 144: 1–35, 2008.

M. Ammann, S. Odoni, and D. Oesch. An alternative three-factor model for international markets: Evidence from the European Monetary Union. *Journal of Banking and Finance*, 36:1857–1864, 2012.

S. Artmann, P. Finter, and A. Kempf. Determinants of expected stock returns: large sample evidence from the German market. *Journal of Business Finance and Accounting*, 39:758–784, 2012.

D. Avramov and T. Chordia. Asset pricing models and financial market anomalies. *Review of Financial Studies*, 19: 1001–1040, 2006.

R. Ball and R. Watts. Some additional evidence on survival biases. *Journal of Finance*, 34:197–206, 1979.

M. Bank, M. Larch, and G. Peter. Investor's compensation of illiquidity - Evidence from the German stock market. *International Journal of Economic Research*, 9:381–408, 2012.

R. Bauer, M. Cosemans, and P. Schotman. Conditional asset pricing and stock market anomalies in Europe. *European Financial Management*, 54:75–89, 2008.

W. Bauman, C. Conover, and R. Miller. Growth versus value and large-cap versus small-cap stocks in international markets. *Financial Analyst Journal*, 54:75–89, 1998.

G. Bekaert, C. Harvey, and C. Lundblad. Liquidity and expected returns: Lessons from emerging markets. *Review of Financial Studies*, 20:1783–1831, 2007.

G. Bekaert, C. Harvey, C. Lundblad, and S. Siegel. What segments equity markets. *Review of Financial Studies*, 24: 3841–3890, 2012.

R. Bennin. Error rates in CRSP and COMPUSTAT: A second look. *Journal of Finance*, 35:1267–1271, 1980.

J. Bley and M. Saad. Idiosyncratic risk and expected returns in frontier markets: Evidence from GCC. *Journal of International Financial Markets, Institutions and Money*, 22:538–554, 2012.

M. Borys and P. Zemcik. Size and value effects in Visegrad countries. *Emerging Markets Finance and Trade*, 47: 50–68, 2011.

M. Brennan, T. Chordia, and A. Subrahmanyam. Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. *Journal of Financial Economics*, 49:345–373, 1998.

A. Brown, D. Y. Du, S. G. Rhee, and L. Zhang. The returns to value and momentum in Asian markets. *Emerging Markets Review*, 9:79–88, 2008.

R. Brückner. Important characteristics, weaknesses and errors in German equity data from Thomson Reuters Datastream and their implications for the size effect. *SSRN Working Paper*, 2013.

R. Brückner, P. Lehmann, and R. Stehle. In Germany the CAPM is alive and well. *SSRN Working Paper*, 2012.

G. Bulkley and V. Nawosah. Can the cross-sectional variation in expected stock returns explain momentum? *Journal of Financial and Quantitative Analysis*, 44:777–794, 2009.

H. Burghof and F. Prothmann. The 52-week high strategy and information uncertainty. *Financial Markets and Portfolio Management*, 25:345–378, 2011.

N. Cakici, F. Fabozzi, and S. Tan. Size, value, and momentum in emerging markets stock returns. *Emerging Markets Review*, 16:46–65, 2013.

L.K.C. Chan, N. Jegadeesh, and J. Lakonishok. Evaluating the performance of value versus glamour stocks. the impact of selection bias. *Journal of Financial Economics*, 38:269–296, 1995.

L.K.C. Chan, J. Karceski, and J. Lakonishok. The risk and return from factors. *Journal of Financial and Quantitative Analysis*, 33:159–188, 1998.

Andy C.W. Chui, Sheridan Titman, and K.C. John Wei. Momentum, legal systems and ownership structure: An analysis of Asian stock markets. *Working Paper*, 2003.

Andy C.W. Chui, Sheridan Titman, and K.C. John Wei. Individualism and momentum around the world. *Journal of Finance*, 65:361–392, 2010.

S. M. Courtenay and S. B. Keller. Errors in databases revisited: An examination of the CRSP shares-outstanding data. *The Accounting Review*, 69:285–291, 1994.

K. Daniel, S. Titman, and K. Wei. Explaining the cross-section of stock returns in Japan: Factors or characteristics. *Journal of Finance*, 56:743–766, 2001.

J. Davis. The cross-section of stock returns and survivorship bias: Evidence from delisted stocks. *Quarterly Review of Economics and Finance*, 56:365–375, 1996.

J. Davis, E. Fama, and K. French. Characteristics, covariances, and average returns: 1929-1997. *Journal of Finance*, 55:389–406, 2000.

W. De Groot, J. Huij, and W. Zhou. Another look at trading costs and short-term reversal profits. *Journal of Banking and Finance*, 36:371–382, 2012a.

W. De Groot, J. Pang, and L. Swinkels. The cross-section of stock returns in frontier emerging markets. *Journal of Empirical Finance*, 19:796–818, 2012b.

L. De Moor and P. Sercu. The smallest firm effect: An international study. *Journal of International Money and Finance*, 32:129–155, 2013.

M. Dempsey. The book-to-market equity ratio as a proxy for risk: evidence from Australian markets. *Australian Journal of Management*, 35:7–21, 2010.

G. Diacogiannis and D. Kyriazis. Testing the performance of value strategies in the athens stock exchange. *Applied Financial Economics*, 17:1511–1528, 2007.

M. Drew, A. Marsden, and M. Veeraraghavan. Small firm effect, liquidity and security returns: Australian evidence. *Journal of Emerging Market Finance*, 5:135–149, 2006.

The Economist. Data birth. fifty years after the dawn of empirical financial economics, is anyone the wiser? 2010.

E. Eiling, B. Gerard, P. Hillion, and F. de Roon. International portfolio diversification: Currency, industry and country effects revisted. *Journal of International Money and Finance*, 31:1249–1278, 2012.

A. Eisdorfer. Delisted firms and momentum profits. *Journal of Financial Markets*, 11:160–179, 2008.

C. Eun and W. Huang. Asset pricing in China's domestic stock markets: Is there a logic? *Pacific-Basin Finance Journal*, 15:452 480, 2007.

E.F. Fama. My life in finance. *Annual Review of Financial Economics*, 3:1–15, 2011.

E.F. Fama and K.R. French. The cross-section of expected stock returns. *Journal of Finance*, 47:427–465, 1992.

E.F. Fama and K.R. French. Differences in the risks and returns of NYSE and NASD stocks. *Financial Analyst Journal*, 49:37–41, 1993.

E.F. Fama and K.R. French. The CAPM is wanted, dead or alive. *Journal of Finance*, 51:1947 – 1958, 1996.

E.F. Fama and K.R. French. Value versus growth: The international evidence. *Journal of Finance*, 53:1975–1999, 1998.

E.F. Fama and K.R. French. Size, value, and momentum in international stock returns. *Journal of Financial Economics*, 105:457–472, 2012.

J. Fletcher. An examination of alternative factor models in UK stock returns. *Review of Quantitative Finance and Accounting,*, 16:117130, 2001.

M. Glaser and M. Weber. Momentum and turnover: Evidence from the German stock market. *Schmalenbach Business Review*, 55:108–135, 2003.

W. N. Goetzmann, R. G. Ibbotson, and L. Peng. A new historical database for the NYSE 1815 to 1925: Performance and predictability. *Journal of Financial Markets*, 4:1–32, 2001.

R. Y. Goyenko, C. W. Holden, and C. A. Trzcinka. Do liquidity measures measure liquidity? *Journal of Financial Economics*, 92:153–181, 2009.

A. Gregory, R. Tharyan, and A. Huang. The Fama-French and momentum portfolios and factors in the UK. *Paper No 09/05, University of Exeter*, 2009.

A. Gregory, R. Tharyan, and A. Christidis. Constructing and testing alternative versions of the Fama-French and Carhart models in the UK. *Journal of Business Finance and Accounting*, 40:172 – 214, 2013.

J. M. Griffin. Are the Fama-French factors global or country specific? *Review of Financial Studies*, 15:783 – 803, 2002.

H. Guo and R. Savickas. Relation between time-series and cross-sectional effects of idiosyncratic variance on stock returns. *Journal of Banking and Finance*, 34:1637–1649, 2010.

R.C. Gutierrez and E. K. Kelley. The long-lasting momentum in weekly returns. *Journal of Finance*, 63:415447, 2008.

M. Hanauer, C. Kaserer, and M. S. Rapp. Risikofaktoren und Multifaktorenmodelle für den deutschen Aktienmarkt. *Betriebswirtschaftliche Forschung und Praxis*, forthcoming.

S. Heston, G. Rouwenhorst, and R. Wessels. The role of beta and size in the cross-section of European stock returns. *European Financial Management*, 5:9–27, 1999.

Joel L. Horowitz, Tim Loughran, and N.E. Savin. Three analyses of the firm size premium. *Journal of Empirical Finance*, 7:143–153, 2000.

K. Hou, G.A. Karolyi, and B. Kho. What factors drive global stock returns? *Review of Financial Studies*, 24: 2527–2574, 2011.

Y. Huang, J. Yang, and Y. Zhang. Value premium in the Chinese stock market: free lunch or paid lunch? *Applied Financial Economics*, 23:315324, 2012.

O. S. Ince and R. B. Porter. Individual equity return data from Thomson Datastream: Handle with care! *Journal of Financial Research*, 29:463–479, 2006.

R. Jagannathan and Z. Wang. The conditional CAPM and the cross-section of expected returns. *Journal of Finance*, 51:3–53, 1996.

S. Kim, D. Kim, and H. Shin. Evaluating asset pricing models in the Korean stock market. *Pacific-Basin Finance Journal*, 20:198–227, 2012.

S.P. Kothari, J. Shanken, and R. Sloan. Another look at the cross-section of expected returns. *Journal of Finance*, 50: 185–224, 1995.

E. Lee, W. Liu, and N. Strong. UK evidence on the characteristics versus covariance debate. *European Financial Management*, 13:742756, 2007.

J. F. LHer, T. Masmoudi, and J. M. Suret. Evidence to support the four-factor pricing model from the Canadian stock market. *Journal of International Financial Markets, Institutions and Money*, 14:313–328, 2004.

J. Lischewski and S. Voronkova. Size, value and liquidity. Do they really matter on an emerging stock market? *Emerging Markets Review*, 13:8 – 25, 2012.

W. Liu. A liquidity-augmented capital asset pricing model. *Journal of Financial Economics*, 82:631–671, 2006.

A. Lo and A. C. MacKinlay. When are contrarian profits due to stock market overreaction? *Review of Financial Studies*, 3:175–205, 1990.

T. Loughran. NYSE vs NASDAQ returns. market microstructure or the poor performance of initial public offerings? *Journal of Financial Economics*, 33:241 – 260, 1993.

A. C. MacKinlay. Multifactor models do not explain deviations from the CAPM. *Journal of Financial Economics*, 38:3–28, 1995.

R. McElreath and C. Wiggins. Using the compustat tapes in financial research: problems and solutions. *Financial Analysts Journal*, 40:71–76, 1984.

T. H. McInish, D. K. Ding, C. S. Pyun, and U. Wongchoti. Short-horizon contrarian and momentum strategies in Asian markets: An integrated analysis. *International Review of Financial Analysis*, 17:312–329, 2008.

S. Nagel. Accounting information free of selection bias: A new UK database 1953-1999. *Working Paper, London Business School*, 2001.

A. Naranjo and B. Porter. Risk factor and industry effects in the cross-country comovement of momentum strategy. *Journal of International Money and Finance*, 29:275–299, 2010.

Gilbert V. Narteaa, Bert D. Warda, and Lee J. Yao. Idiosyncratic volatility and cross-sectional stock returns in South-east Asian stock markets. *Accounting and Finance*, 51:1031–1054, 2011.

T. Nijman, L. Swinkels, and M. Verbeek. Do countries or industries explain momentum in Europe? *Journal of Empirical Finance*, 11:461–481, 2004.

Marc Reinganum. Market microstructure and asset pricing: An empirical investigation of NYSE and NASDAQ securities. *Journal of Financial Economics*, 28:127–147, 1990.

B. Rosenberg and M. Houglet. Error rates in CRSP and Compustat data bases and their implications. *Journal of Finance*, 29:1303–1310, 1974.

F. Rossi. U.K. cross-sectional equity data: do not trust the dataset! the case for robust investability filters. *Working Paper, UCD Michael Smurfit Graduate Business School*, 2011.

G. Rouwenhorst. Local factors and turnover in emerging markets. *Journal of Finance*, 54:1439–1464, 1999.

P. S. Schmidt, U. von Arx, A. Schrimpf, A. F. Wagner, and A. Ziegler. On the construction of common size, value and momentum factors in international stock markets: A guide with applications. *Working Paper 11/141 Center of Economic Research at ETH Zurich*, 2011.

A. Schulz and R. Stehle. Buchwert-Marktwert-Verhaeltnis, Size und Beta als Erkärungsvariable für die Renditen deutscher Aktien. *Working Paper, Humboldt-Universität Berlin*, 2002.

T. Shumway. The delisting bias in CRSP data. *Journal of Finance*, 52:327–340, 1997.

T. Shumway and V.A. Warther. The delisting bias in CRSP's Nasdaq data and its implications for the size effect. *Journal of Finance*, 54:2361–2379, 1999.

A. Siganos. Can small investors exploit the momentum effect? *Financial Markets and Portfolio Management*, 24: 171–192, 2010.

I. Soumare, E. K. Amenounve, O. Diop, D. Meite, and Y. D. N'sougan. Applying the CAPM and the Fama-French models to the BRVM stock market. *Applied Financial Economics*, 23:275–285, 2012.

N. Ulbricht and C. Weiner. Worldscope meets Compustat: A comparison of financial databases. *sbr-Working Paper*, 2005.

M. Umutlu, M. Akdeniz, and A. Altay-Salih. The degree of financial liberalization and aggregated stock-return volatility in emerging markets. *Journal of Banking and Finance*, 34:509–521, 2010.

J. van der Hart, E. Slagter, and D. van Dijk. Stock selection strategies in emerging markets. *Journal of Empirical Finance*, 10:105 – 132, 2003.

J. van der Hart, G. de Zwart, and D. van Dijk. The success of stock selection strategies in emerging markets: Is it risk or behavioral bias? *Emerging Markets Review*, 6:238 – 262, 2005.

X. Wang. Size effect, book-to-market effect, and survival. *Journal of Multinational Financial Management*, 10: 257–273, 2000.

A. Waszczuk. The risk–based explanation of return patterns – Evidence from the Polish Stock Market. *Emerging Markets Review*, 15:186–210, 2013.

M. Weinstein. Larry fisher: ourSherpa into the mountains of data. *Review of Quantitative Finance and Accounting*, 35:127–135, 2010.

Y. Yao. Momentum, contrarian, and the January seasonality. *Journal of Banking and Finance*, 36:2757–2769, 2012.

X. Zhang. Specification test of international asset pricing models. *Journal of International Money and Finance*, 25: 275–307, 2006.